Reflections in Toxicology

Morris F. Cranmer, PhD.

Director, National Center for Toxicological Research, Jefferson, Arkansas 72079

Thank you for your kind introduction, Doctor Friess. I will try to weave a few of my own comments and opinions with comments by previous speakers and observers into a review of what has been stated or implied. I will also try to document what research the NCTR is doing that may impact on these problem areas. I have borrowed a few slides from previous speakers to emphasize certain points.

Few people will dispute the fact that chemical technology has in large measure contributed to the achievement of our present standard of living. Accompanying these benefits, however, are the many subtle and sometimes gross effects that are threatening the health of our society. The existing implications on future generations demand the adoption of a rational policy on chemical utilization that will enable the highest possible standard of living accompanied by an acceptable risk-to-benefit ratio.

Persons suffering from an incurable, fatal disease would not want to have treatment with a drug withheld because of a potential danger of developing cancer far into the future. Similarly, persons beyond the reproductive years certainly have less concern for exposures to chemicals that produce birth defects or genetic change than do young adults. In short, society accepts considerable risks when the risks are necessary and when acceptable alternatives do not exist, but is predictably unwilling to accept risks when the information quantitating those risks is not available.

Efficacious products are generally approved for use if there is an acceptable safety margin between anticipated residues, by appropriate usage patterns, and that level estimated to be safe to humans. The toxicologist is faced with the dilemma of estimating risk to an enormous and variable human population from small numbers of highly controlled experimental animals. Thus, there exists a considerable potential for error in assessing the risk/benefit ratio under present conditions.

Several facts which contribute to the uncertainty of toxicological evaluations should be stated clearly. There is no way to guarantee absolute safety! Small populations of experimental subjects, either animal or man, provide an imprecise basis for comparison to a large human population of variable genetic/disease states, cultural backgrounds and ages. Toxicological assessments are made singularly and humans are exposed to a milieu. It should be equally clearly understood that a proper experimental design will minimize noise and maximize comparisons and that we are constantly expanding our toxicological armamentarium. Doctor Rall explored many of the strengths and weaknesses of toxicological evaluation in his paper on Wednesday morning.

Is the toxicologist faced with a paradox of absolutes? What are the approaches available in attempts to generate reasonable policy and guides for chemical use and control? An examination of the involvement of the toxicologist in guaranteeing an adequate and acceptable food supply will be illuminating. There are three major control strategies available for the regulation of toxicants, including carcinogens: 1. the all-or-none approach, (for example, the Delaney Clause of the Food, Drug, and Cosmetic Act [FD&C Act]); 2. the use of safety factors (commonly applied to non-

carcinogenic lesions); and 3. statistical extensions beyond the experimentally observable range. Each approach has its proponents and critics, its advantages and disadvantages. Doctor Schneiderman made several comments on the Mantel-Bryan model for extrapolation, and I will try to expand this point.

All-or-None Approach

The Delaney Clause of the FD&C Act is an all-or-none approach and an understanding of complications in the quest of absolute safety is required. The Delaney Anticancer Clause contains two main segments; one for human and one for animal food additives. The segment addressing human food additives states that in evaluating the safety of such compounds used in food-producing animals, consideration must be given to the safety from possible residues in the products of those animals which are a source of food for man. When there is insufficient evidence to establish that a finite or negligible residue of the compound is safe in human food, or when the anticancer clauses contained in sections 409(c)(3)(A), 512(2)(1)(H), and 706(b)(5)(3) of the Act are applicable, a zero tolerance (no residue) must be required. Under the provisions of the anticancer clauses, no compound may be administered to animals which are raised for food production if such compound has been shown to induce cancer when ingested by man or animal, unless such compound will not adversely affect the animal and no residues, as determined by methods of analysis prescribed or approved by the Secretary (DHEW), are found in the edible products of such animals under conditions of use specified in labeling and reasonably certain to be followed in practice.

How Does One Establish the Toxicity; e.g., Carcinogenesis of a Compound?

A protocol advanced by the National Cancer Institute for carcinogen screening calls for 50 male and 50 female animals to be tested at or near the maximum tolerated dose and a like number at half that dose. A maximum tolerated dose ideally would be that which does not kill the animal except via tumor production in significantly less than a normal lifespan. The choice of using high doses is a statistical expedient in order that high incidences of tumors above background can be detected with small sample sizes. There is no biological basis for use of high doses. Such high doses may completely alter metabolic pathways, absorption and distribution.

Positive and negative results are treated in a completely different manner. If the screening test shows positive carcinogenic action, under the Delaney Clause there is no alternative but to ban the compound even if more than adequate information was available to perform a risk/benefit analysis. All too often, a negative result is interpreted as indicating a noncarcinogenic compound. The negative cannot be proved statistically. Thus, it is common practice to take an arbitrary fraction, say 1/100, of the minimum "no-effect" dosage as safe. Again, the minimum "no-effect" dosage is ill-defined and is a random variable depending on the number of animals tested. Two statements will be repeated in this and subsequent discussions. First, it is argued that such an approach has worked over the years. Second, do we have epidemiological evidence to show that no small increases in cancer have resulted from such environmental chemicals? Until recently we were not aware of the vinyl chloride problem even though billions of pounds are manufactured each year. Most new chemicals have not been in the environment long enough for effects to be noted where long latent periods may exist.

Zero Tolerance—No Residue

If one defines zero as complete absence, the dilemma of a no-residue concept quickly takes form. First, let us consider the ways by which one might attain the complete absence of a residue. The first would be to never allow contact, and the second would be to consider a rate of

removal or transformation that after a given waiting period would result in complete removal.

Since the first approach effectively eliminates the use of a chemical, let us proceed with the concept of removal. The process of removal may be passive, e.g., the removal of a persistent pesticide from a food by rain or washing, or it may be via active excretion or metabolism. If an enzymatic process is involved, the process will be accelerated.

The rate of an enzymatic process can be either zero ($^{dc}/dt = Ko = KoC^{o}$), first order ($-^{dc}/dt = k_1OC = k_1C^{1}$), or second order ($-^{dc}/dt + k_2(C \times C) = k_2C^{2}$). If one solves these equations for time needed to achieve the removal of the last molecule, it is a very long time indeed. Further practical complications arise via the determination of analytical methodologies which would be acceptable for determining zero. To carry the discussion to the extreme, we would have to analyze the complete extract of the complete sample with an analytical sensitivity of one molecule.

We will probably all agree that there are certain advantages to the use of biologically active chemicals and there is also the need to insure that the population is not exposed to hazardous residue levels. One reasonable approach would be to insure the absence of any hazardous quantity of a residue. The advantage of this approach would first be to fix the amount of residue which is expected to be hazardous and at the same time determine the sensitivity of a method required for analysis to support regulatory actions. Rephrased, requirement for methodology would be defined by need rather than state of the art. Use of a chemical would not be approved until acceptable methodology was developed.

As is often the case, problems are not solved, they simply are reshaped, for we now have the problem of determining the acceptable residue level.

The Delaney Clause, or any similar all-or-none approach, is likely to be inadequate in two respects. First, it provides a false sense of security by ignoring the problem of "false negatives" which may result from inadequate testing. The FDA is charged with the responsibility of attempting to minimize such occurrences, but the question remains as to how to best accomplish this formidable task. Second, because of current toxicological ignorance we have little to offer as a substitute for the Delaney Clause which requires banning of food additives shown to be carcinogenic in animal tests. However, with adequate data, yet to be produced, the benefits of a food additive in preventing food poisoning, for example, might be documented to far outweigh a carcinogenic risk which may occasionally occur only late in life.

Safety Factors

Safety evaluation at the present time is founded on the concept of the "Maximum no-effect dose." The procedures are designed to determine the intake over extended periods (including a lifetime) that will not produce the injurious effects characteristic of the substance when given in large, that is, toxic amounts. Also important is the exclusion of the possibility that these "subtoxic" amounts will produce some hitherto unsuspected reaction. A summary of the kinds of specific studies usually undertaken can be found in the paper by Friedman and Spiher (*FDA Papers*, Nov. 1971).

The unique difficulties in safety evaluation arise from the unusual goal of attempting to prove scientifically that no deleterious effect has taken place, i.e., to prove the negative. Experiments are usually designed to establish that phenomena, apparently resulting from experimental manipulations, are real, are not artifacts or have not occurred simply by chance. On the other hand, the more appropriate concern would be to ensure that the *absence* of positive findings (*assuming adequate protocols and procedures*), is not due to chance or to the *inadequacies of sample size*. Pursuing this point supports the awareness that positive findings may be artifacts and

therefore adequate probing of techniques and replication of experiments to verify findings is mandatory. Insistence on any desired degree of assurance against making a wrong conclusion is standard operating procedure. Conventionally, a statistically significant finding must have a probability of no more than one chance in twenty of being a chance occurrence, and often risks of only 1 in 100, or 1 in 1000, or less, are desired. Clearly the severity of an all or none approach to avoid the risk of a *false* positive reinforces the desire of a petitioner for the clearance of a compound. Have we dealt equally with false negatives?

A practical approach for dealing with these uncertainties for noncarcinogens has been the use of the 100-fold margin of safety. Substances to be added to food should not demonstrate an effect in animals when fed at a dose at least 100 times greater than the likely human exposure. Our intuition tells us that this approach has usually worked very well; however, we should not forget the absence of an experimental or theoretical basis. When followed blindly, rather irrational experimental practices, interpretation and rationalization can be made.

There have been attempts to apply safety factors to carcinogens in our food supply. One of the latest discussions was by Carrol S. Weil (1972 *Toxicology and Applied Pharmacology, 21*(4): 454) where a safety factor of 5,000 was suggested. Weil argued, as had Friedman, that it was contrary to "scientific judgment" to try to extrapolate mathematically beyond the range of experimental observation. Weil suggested that it was, however, more scientific to use a safety factor of 5,000.

The application of a safety factor established from a "no effect level" in a toxicological evaluation has a number of pitfalls which were succinctly summarized by Weisburger and Weisburger (1968, *Food Cosmetic Toxicology*, 6: 235–242):

> It seems to us a "no effect dose" for a carcinogen is a highly relative level which applies only for the precise experimental conditions generated. While similar considerations hold for drugs, the risk is not nearly so intense. More often than not, an improper dose rate for rapidly acting drugs is detected almost immediately and appropriate remedial action can be taken. With chemical carcinogens and their long, latent period, the disease condition resulting from inappropriate selection of dose levels and alteration of environmental conditions leading to potentiation may become visible only years after the exposure. At that time remedial action is obsolete and often worthless.

It is necessary to add to the Weisburger remarks that a no-effect dose, with the exception of a threshold, is sample-size dependent and therefore is *not* some absolute reference point.

A number of the contributors and observers, including Doctor Worcester and Doctor Henderson, during the discussion periods on Wednesday and Thursday, spoke of thresholds.

A few comments on "threshold" are an appropriate prelude to a discussion of methods for mathematical extrapolation. The concept of a threshold dose is based on the premise that a smaller dose will not produce an effect. There are several problems with demonstrating the reality of a threshold. More refined methods of observation may lower the observed threshold: repeated examination of the bioassay will demonstrate variability even within the same individual, and heterogeneity of the population will influence the responses observed. Many toxicologists have stated that for any compound there must be a "biologically insignificant dose." There is little doubt that this is true: however, what is our definition of insignificant. A case in point are reports which have been used to estimate that 3–5 percent of those people hospitalized have drug complications severe enought to extend their duration of care, a very ominous statistic.

Mathematical Extrapolation Models

One of the real pleasures of my scientific career has been being associated with discussing mathematical models with Dr. Dave Gaylor, and much I will say is his work. Due to, at least, the toxicological uncertainties of extrapolating risks from relatively high experimental dosages in animals to low human exposure levels,

there are many people who propose complete prohibition when a chemical is demonstrated to be a carcinogen. A modification would be to use a conservative method of linear extrapolation from an upper confidence limit on the experimental result back to a zero response at zero dosage. This procedure is described by Gross, Fitzhugh, and Mantel (1970) and the FDA Advisory Committee on Protocols for Safety Evaluation (1971). This procedure is based on the premise that at low dosages, many dose-response curves are concave upward and a straight line is a conservative upper limit to such curves. In the simplest case with a single dosage and no spontaneous background occurrence of tumors, the extrapolation would proceed from setting an upper confidence limit on the observed tumor rate at the experimental dosage, d, and constructing a line back to zero. Such a straight line is likely to be above the true dose-response curves at low doses. For low dosages, the one-hit curve is approximately proportional to dosage (linear).

For the particular experimental conditions, a conservative upper limit, p_o, can be estimated for any low dosage, d_o. If a threshold dosage does exist below which no tumors are produced, the true tumor rate at d_o may be zero. An objection to this method of linear extrapolation is that in order to obtain small risk levels, p_o, the levels of d_o which could be tolerated often would be too small to make the food additive effective for its intended purpose. However, this procedure encourages better experimentation in that as the number of animals tested is increased, the upper confidence limit will generally decrease thereby increasing d_o for any given level of estimated risk, p_o. The more complicated and common situations of non-zero spontaneous background and multiple dosages are discussed by Gross, Fitzhugh, and Mantel (1970).

Of the common mathematical models often proposed for extrapolation (one-hit, logistic, extreme value, and probit) the Mantel-Bryan (1961) procedure proposes the use of the probit.

The model for extrapolation usually cannot be determined from experimental results. For example, consider the probit, logistic, and one-hit curves which all give a 50% tumor response at a unit dose and 16% tumor response at 1/4 that dose. These curves would be indistinguishable in the 8% to 92% tumor response range as usually observed in experiments. Several thousand animals would be required to distinguish between the probit and logistic curves in the 2% to 4% response range with no guarantee that either model would be applicable at lower levels. In Table 1 are shown extrapolated doses producing small risks where the experimental data appear almost identical in the 8% to 92% tumor response range (FDA Advisory Committee on Protocols for Safety Evaluation (1971)).

Table 1.—Doses required to give low estimated risks from experimentally indistinguishable results with 8–92% tumors (a dose of one unit produces 50% tumors).

Estimated Risk	Probit	Logistic	One-hit
10^{-3}	1.5×10^{-2}	3.1×10^{-3}	1.4×10^{-3}
10^{-6}	1.4×10^{-3}	9.8×10^{-6}	1.4×10^{-6}
10^{-8}	4.1×10^{-4}	1.6×10^{-7}	1.4×10^{-8}

For example, if a dose of one unit produced 50% tumors, then a dose of .015 units would be expected to produce 1 tumor in 1000 animals, assuming extrapolating with the probit curve. Extreme differences between models in estimated doses are noted when extrapolating to a 1 in a million risk. The "extreme value" curve, another possible model, would generally lie between the probit and logistic, depending on slopes, Chand and Hoel (1973). Thus, the choice of a model for extrapolation is extremely critical, the one-hit being the most conservative and the probit the least conservative of those examined here.

Fig. 1 illustrates procedures utilized in the Mantel-Bryan model. The Mantel-Bryan (1961) procedure utilizes a linear relationship between probits and log dosage. They propose a conservative slope of one probit per 10 fold reduction

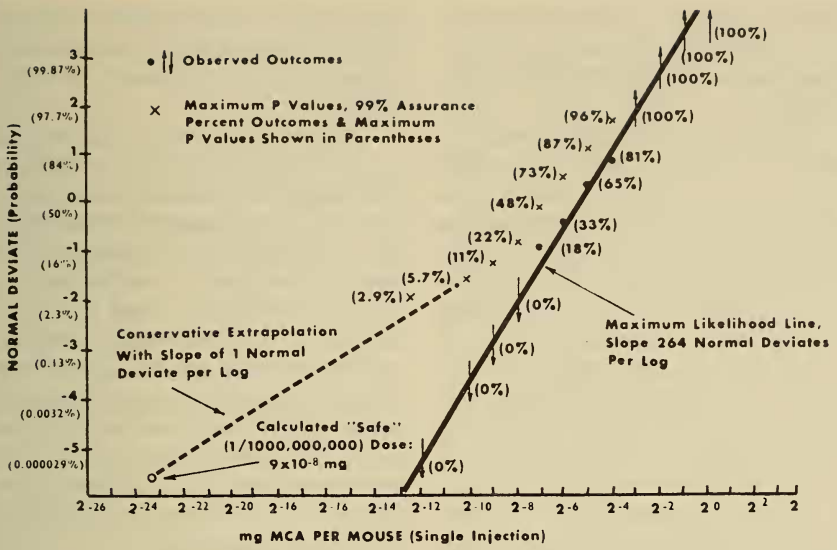

Fig. 1.—Estimation of the "safe" dose from test results with a carcinogen, methylcholanthrene, at several dose levels.

in dose. These lines are fitted at moderate to high responses, usually high experimental doses and generally using homogeneous groups of animals, which would be expected to produce steep slopes. There is no guarantee that slopes might not be less than one at low doses to which a heterogeneous human population is exposed. In fact, the dose-response in the smoking lung cancer data for man (percent of men developing lung cancer versus number of cigarettes smoked per day) gives a probit slope of about 0.75. However, a slope of one, hopefully, represents a conservative slope in the dosage range below the experimental dosages. The only notable exceptions with experimental slopes less than one which have been established are the hormonal animal feed additive growth promoters. In such cases, a slope less than one would be recommended for extrapolation.

The Mantel-Bryan procedure has these advantages: it does not require an experimental estimate of the slope; it does not require the demonstration of a statistically significant increase in tumors (which depends heavily upon the number of animals tested); it allows for a non-zero spontaneous background tumor rate; however, more research is needed where background rates are high, often resulting in treated animals with fewer tumors; it considers multiple dosage experiments. The estimated risks using the Mantel-Bryan procedure depend upon the degree of the uncertainty in the experimental data by starting the extrapolation from upper confidence limits on tumor rates and not upon proof of carcinogenicity. It does not assume that the dose-response relationship is probit log-dosage at low dosages. If a non-probit response curve is plotted on a probit scale and if its derivative is always greater than one, then the slope of one applied by the Bryan-Mantel procedure at low dosages gives a proportion of tumors which is higher than the actual proportion. However, if the same principle is applied to logistic plotting, lower extrapolated dosages will result for a given risk. It is not necessary to extrapolate to a "virtually safe" safe risk of 1 in 100 million. This value was selected by Mantel and Bryan as an "illustrative" value which probably would not be in conflict with the intent of the Delaney Clause. "Acceptable risk" is a social judgment which will have to be made by open discussions after weighing the bene-

fit of each chemical, its possible synergism with other compounds, and the uncertainty in extrapolating from animals to man.

A dichotomous procedure could be employed by extrapolating with an extremely conservative linear model for experimentally demonstrated carcinogens and extrapolating with a less conservative procedure, such as the Mantel-Bryan procedure for chemicals not demonstrated to be carcinogens. A dichotomous extrapolation rule may not encourage good experimentation to detect tumors in order that a less conservative extrapolation procedure could be used leading to higher tolerance levels.

The extreme differences between models for extrapolating to low risks have been demonstrated in Table 1. Even given a particular model, e.g. the probit, the slope was used for extrapolation produces widely different results (Table 2) where extrapolations were

Table 2.—Fraction of experimental dose using probit extrapolation with different slopes for an estimated risk of 1 in 100 million.

Observed tumors	Slope = 1	Slope = 1.5	Slope = 2.0
0/50	1/18,000	1/690	1/135
0/100	1/8,300	1/410	1/91
0/500	1/1,800	1/150	1/42
0/1,000	1/1,000	1/100	1/32

made from the upper 99% confidence limit.

For example, if no tumors were observed in 100 animals, one could be 99% confident that the true risk is no more than 1 in 100 million if the dose-response curve has a slope of one probit per factor of 10 change in dose, when the experimental dosage is divided by 8300. In Table 2 it is illustrated that the current practice of taking 1/100 of an observed no-effect level for 100 animals would provide a risk of approximately 1 in 100 million if the response curve were a probit with a slope of two. However, if the slope were actually 1, then the estimated dose should be about 90 times smaller. Thus, not only is the choice of an extrapolation model critical, but the parameters used in the model, particularly the slope, are critical.

Unfortunately, one cannot verify experimentally the correct curve (model and slope) to use for extrapolation at extremely low dosages. It would be useful to obtain dose-response curves at levels lower than currently used in experiments. Perhaps a data bank can be accumulated for low dosage levels which provoke few, if any, tumors. Such data might eventually provide reasonable estimates of low dosage exposures, or perhaps, a check on the form of mathematical models. One difficulty would be differences in protocol employed by different investigators.

As was discussed by Doctors Schneiderman, Rall and Newill, we are still faced with the uncertainties in extrapolating from well-controlled animal experiments to heterogeneous human populations. Thus, there are those who rightly contend that no method of precise mathematical extrapolation exists to date. However, as I mentioned before, using 1/100 of an observed no-effect level as relatively safe is, in fact, performing an arbitrary and crude extrapolation which ignores the uncertainty in the experimental data. Predictions of tolerable dosages from animal experiments must be made. Should these predictions be made with or without the benefit of all of the scientific knowledge at hand? It is interesting to compare the Mantel-Bryan procedure with the 1/100th rule. As seen in Table 2, Mantel-Bryan would set lower levels using a slope of one if a risk of 1 in 100 million is used. Adopting an extrapolation slope of 1.5 would be in agreement with the 1/100th rule for large experiments.

An important aspect of extrapolation is the choice of the dose scale. Log dosage on a per body weight basis is frequently used as is ppm. Dosage on a surface area basis as mentioned by Doctor Rall has been investigated and appears to give a better fit to experimental data in some cases and appears useful when extra-

polating from small to large animals. No single choice can be recommended. The tendency is to express dosage in terms that give a nearly linear fit to the data in the experimental range.

Data in man, either dose-response or metabolic, may suggest greater or lesser sensitivity than the experimental animal. Human data seldom is available, but when it is available it generally is not clear how such data should be employed in a mathematical procedure for prediction of dosages producing low risks. Much more epidemiological data is needed. A current example of this is the need to use the human data from benzidine as a component of setting water effluent standards by the EPA.

Petitioners should be encouraged to conduct experiments in more than one species. Selecting the lowest tolerance for extrapolation to man from the species tested, in order to be conservative, may tend to discourage testing in several species. This appears to be the most prudent approach. Perhaps to encourage testing in more species, the slope for extrapolation could be increased as the number of species is increased. For example, if the Mantel-Bryan procedure is used in a single species, a slope of one could be used unless experimental results indicated a shallower slope. If more species were tested, steeper slopes could be allowed for extrapolation with each species, still employing the lowest tolerance from among the species tested if the experiments were done with sufficient precision that the lower confidence of the slope could be determined statistically with high confidence to be greater than the slope to be used. For example, an experimental slope of 4 with a lower boundary of 3 might allow for using 1.5 rather than 1. This procedure is only a suggestion which should be investigated with existing data to determine its workability.

Another important aspect of extrapolation is determining the level of an acceptable risk. This is a social-political decision which cannot be made by the scientist alone, but requires a risk-benefit analysis with input by many segments of society. This is an awesome task, but we are faced with it daily in setting speed limits, building codes, etc. Admittedly, it may be easier to perform risk-benefit analyses for many aspects of life than for food additives. Few, if any persons, would want a potential carcinogen added to the food supply if its only benefit were esthetic. There would be no need for extrapolation to a tolerable dosage and the Delaney Clause should remain unchanged. However, if an additive is a preservative which prevents or lessens the risks of other diseases, a nutrient which may reduce the risk of other diseases, or improve nutrition by making food less expensive, then a risk-benefit analysis may be in order. For comparative purposes, dose reduction factors for a risk of 1 in a million are given in Table 3

Table 3.—Fraction of experimental dose using probit extrapolation with a slope of one for an estimated risk of one in a million.

Observed tumors	Fraction of experimental dosage
0/50	1/2,500
0/100	1/1,140
0/500	1/250
0/1,000	1/140

using a probit slope of 1. Again, with a large number of animals, the 1/100th dosage of an experimental no-effect level is not too different.

If the extrapolations were correct what does a risk of 1 in 100 million for a lethal tumor really mean? Approximately 1/6 of the people in the United States eventually die due to cancer. An additional 1 in a 100 million would be unnoticeable. Mantel and Bryan suggested that a calculated risk of 1 in 100 million is the practical equivalent of 0 since the conservative procedure used, if correct, sets 1 in 100 million as the *upper* limit on the true risk. The Mantel-Bryan procedure does not attempt to accept a risk of 1 in 100 million but is directed toward a zero risk not exceeding 1 in 100 million.

Table 4.—Probability of tumor incidence estimated using Mantel and Bryan.

Benzo (a) pyrene mg/kg/man/day	Probit slope 1.0	Probit slope 1.5
.010	1×10^{-6}	1×10^{-14}
.020	1×10^{-5}	1×10^{-12}
.040	2×10^{-5}	1×10^{-11}

Some estimated risks are calculated by Friedman (1973) (Table 4) using the Mantel-Bryan procedure based on a mouse intubation study for benzo(a) pyrene by Berenblum and Haran (1955). Depending on daily intake at human exposure levels, estimated risks range from 2×10^{-5} to 10^{-6}, depending heavily upon the slope used. These data illustrate that we already may be accepting what some people would regard as a relatively high risk, 2 in 100,000, in our food supply from charcoal broiled meat. Of course, the individual can make a choice in this instance. Such information, which is often meager and tentative, may not be available or meaningful.

Experimental Design

In testing for carcinogenicity, it is not clear that current experimental designs and methods of analysis are the best that can be developed. It is difficult to detect and estimate the dose for even a high risk, say .01, when the spontaneous background rate is high, say 0.10. However, it may not be desirable to choose a strain of a species of animals with a 0 or near-0 spontaneous rate, as that strain may be resistant to the chemical. It may be desirable to consider relative rather than absolute rates.

The choice of responses to analyze (e.g., proportion of animals with tumors, number of tumors per animal, or time to tumor) will dictate the experimental design. Consideration must be given to the range of dosages, number of dosages, number of animals, length of feeding (total dose), and times of sacrifice, if any.

If a procedure such as the Mantel-Bryan procedure were adopted for extrapolation, it is possible to calculate "acceptable dosages" for given risks as a function of the proportion of the experimental animals producing tumors (which may be zero) and the number of animals employed.

Considerably more research is needed in the development of experimental protocol for predicting carcinogenicity of chemicals, and I feel the NCTR will impact heavily on this area.

Time to Tumor Occurence

A risk of 1 in 100 million represents an additional two people, now living in the United States, who would eventually die of cancer rather than due to some other cause. This raises an important and difficult question. Since cancer generally occurs late in life, would these two cases result in the loss of life of a few days, weeks, or years? Murray and Axtell (1973) investigated this question. The question of time to tumor occurrence becomes critical. Extrapolating to low dosages from a classical dose-response curve does not consider the time at which tumors occur. The experimental response may be the gross percentage of tumors occurring over the lifetime or up until the termination date of an experiment. One recognized difficulty with gross rates is that the animals at the highest doses may have shorter life spans due to toxicity of the chemical and therefore have less opportunity to develop tumors. Thus, the highest dosage may exhibit the lowest gross proportion of tumors. Some, but not all, researchers have made adjustments for changes in mortality due to competing causes of death. The simplest device which has been employed is to sacrifice animals at a fixed point in time (e.g., 18 months with mice) and to observe the percentages of those animals possessing particular tumors. This gives the proportion with tumors to a given time which can be plotted against dose. As an attempt to study time to tumor development, serial sacrifice experiments have been employed with scheduled sacrifices at several points during the life span of the animals. Such

experiments, require a large number of animals. A procedure such as Mantel-Bryan could be used for extrapolation at each time of sacrifice. Now a difficult question arises, does one use the highest tolerable dose found at different sacrifice times or should more weight be given to the earlier tumors? If more weight should be given to early tumors, how much? Also, a great deal of information is generally lost due to animals that die before the termination date or between serial sacrifice times of an experiment. In fact, it may be these animals which die early that contain the most important information because they afford an opportunity to observe tumors early in life.

Time Concept

I hope the highlighting of the work of Blum, at the beginning of this discussion, indicates that we in chemical toxicology are not completely ignorant of the contributions made in radiation as was suggested by one observer.

Blum, in his 1959 work on "Carcinogenesis by Ultraviolet Light", demonstrated a log normal distribution of exposure and cancer development time. (Princeton University Press, Princeton, New Jersey (1959)). It was Druckrey however who dramatized the relationships of time, dose, and dose rate for chemicals in a 1967 monograph representing 25 years work and 10,000 experimental animals. (Quantitative Aspects in Chemical Carcinogenesis, U.I.C.C. Monograph No. 7, Potential Carcinogenic Hazards from Drugs, pp. 60–77 (1967)).

Fig. 2 was chosen because it demonstrates that over a considerable dose range there is not an experimentally observable no effect or threshold dose for diethylnitrosamine. Mr. Wands asked questions of risk benefit and used nitrosamines as an example. Also this figure is the work of Druckrey on which much of which I will speak is based. Can we further quantitate risk?

Druckrey emphasized that if proper scientific judgments are to be made re-

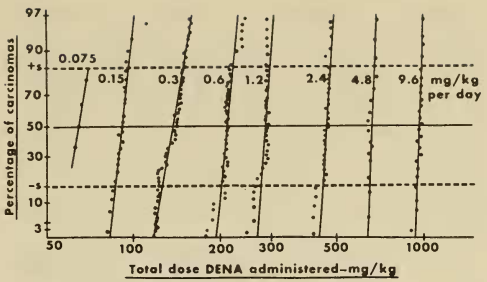

Fig. 2.—Dose-response relationships for the carcinogenic action of diethylnitrosamine (DENA) in BD II rats.

garding risks from carcinogens, knowledge of the pharmacological relationships, especially dose-response relationships, must be established. He pointed out that thorough investigations of these fundamental problems are only possible in systematic, highly controlled, mutually comparable animal experiments. He further pointed out that true advances can only be expected from quantitative results that are expressible in measurement and number and are available for criticism.

Druckrey first demonstrated with 4-dimethyl-amino-benzene the dependancy of time and dose. According to the relationship developed, the product of daily dosage and induction time is a constant ($k = dt$). He observed that the same carcinogenic response was obtained for smaller dose rates and total doses if time was extended ($k = dt^n$). These results, and others, suggested that primary carcinogenic effects for 4-DAB remain irreversible over a whole lifespan and suggested the appropriateness of the concept of "*Summation Action*" which he had introduced 20 years earlier.

Druckrey extended his observations to 4-dimethyl-amino-stilbene (DAST) as illustrated in Fig. 3.

The plot of percentages of tumors expressed as probits vs. log total dose of 4-dimethyl-amino-stilbene (Druckrey, Schmahl, and Dischler, 1963), demonstrates a parallel and linear relationship between probits of ear duct and mammary carcinoma and log total dose. This clearly demonstrates that for 4-DAST

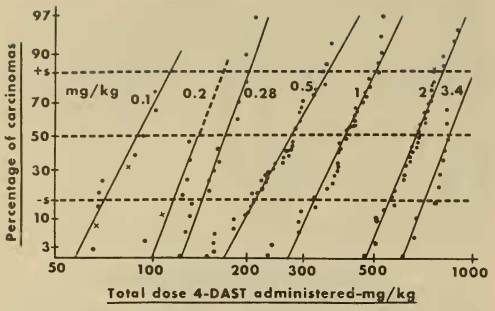

Fig. 3.—Incidence of carcinomas is dependent on the sum of doses, 4-dimethylamino-stilbene.

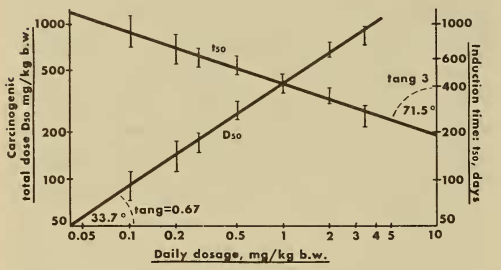

Fig. 4.—Linear dependence of the median carcinogenic total dose D_{50} and of the median induction time T_{50} on the daily dosage of 4-dimethylamino-stilbene.

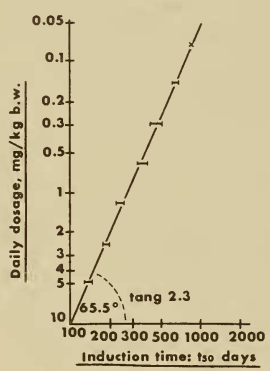

Fig. 5.—Linear dependency of the median induction times (T_{50}) upon the daily dosages of diethylnitrosamine.

the total dose required to produce a given incidence is clearly smaller at smaller dose rates.

A replot of log total dose, and median induction time, vs. log dose rate also demonstrates a linear relationship between the log median induction time and log dose rate (Fig. 4).

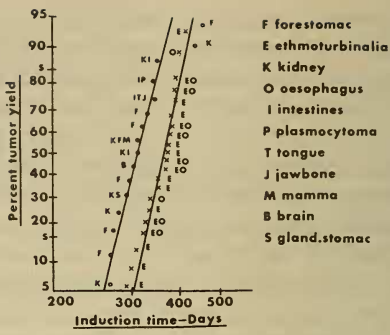

Fig. 6.—Normal distribution of the induction times in carcinogenesis, dependent upon the variety of organs of tumor development in BD rats.

This plot identifies the limiting component lifespan. This figure illustrates for a replot of Druckrey's data, that time begins to be limiting as an experimental variable, a fact that toxicologists and statisticians must work together to exploit. Doctor Schneiderman's comments of a 300 year component in cigarette smoking however demonstrates the complexity of the concept. The difficulty in observing this is not great but attention should be placed on the lack of evidence or suggestion of a sub-threshold dose (Fig. 5). For this specific example, $n = 3$ and $k = dt^3$.

Surprising to me, but as you can see from Fig. 6, the linear relationship held for a number of organ and tumor types generated with methyl nitrosourea and n-nitroso piperidine.

What data exists for man? Doll advanced the concepts relating time and dose still further in his 1970 paper read before the Royal Statistical Society (The Age Distribution of Cancer: Implications for Models of Carcinogenesis. *J. Royal Stat. Soc.* 134 (2): 133–166 (1971)), where he proposed the $I_t = b\ (t\text{-}v\text{-}w)^k$ where v is time of beginning exposure, and w is the minimum time for clinical recognition for a tumor. Doll applied this concept to data on the incidence of bronchial carcinoma in cigarette smokers.

Doll examined data reported by Day in 1967 on tumors from tar painted on the skin of mice. The n obtained for cigarette smoking in man and skin tumors in mice were strikingly similar and encouraging to those that are supportive of efforts to describe mathematically, similarities in cancer responses.

Doll points out that a wide range of pairs of values for k and w in the formula $I_t = b(t\text{-}v\text{-}w)^k$ would fit the data and that a better approach would be to design protocols to estimate the values independently. This is an important point that the carcinogenesis protocol at NCTR addresses. Doll further pointed out that examination of skin cancer generated by benzo (a) pyrene would fit $b\alpha$ $dose^2$ better than $b\alpha$ dose.

There has been some more recent research in statistical techniques to analyze time to tumor data. One useful measure of the impact of carcinogenesis on a population may be age-specific incidence rates or the amount of life shortening due to a tumor. Time to tumor data requires life time studies to estimate a relationship between tumor rates, time, and dosage. This is followed by extrapolations to low dosages which may project time to tumor development beyond the normal lifespan of an animal. Then, for a given dosage a time to tumor distribution must be employed to estimate the proportion of animals expected to develop tumors within their lifespan before dying of other causes. The approach is quite complicated and depends on mathematical assumptions which need to be checked experimentally (the statistical distribution of time to tumors and their relationships with dosage) for many different types of chemicals, tumors, and experimental animals. Such experiments require survival studies and several dosages employing large numbers of animals. Again, there does not appear to be sufficient evidence at this time to recommend specific procedures. It is important to obtain dosage rate effects on the time pattern of response, especially to determine the extent it influences incidence rate and to the extent it influences time to tumors at all incidence levels.

Albert and Altshuler (1973) have developed a mathematical model for predicting tumor incidence and life shortening based on the work of Blum (1959) on skin tumor response with chronic ultraviolet irradiation in mice and on Druckrey (1967) for a variety of chemical carcinogens in rodents. Albert and Altshuler have investigated radiation cancer in mice exposed to radium and also to cigarette smokers. In review, the basic relationship used is: $d t^n = c$, where d is dosage, t is the median time to occurrence of tumors, n is a parameter greater than one, and c is a constant depending on the given experimental conditions. It is of interest to determine the time it takes for a small proportion of the population to develop tumors. With this formulation, as the dosage is increased, the time to tumor occurrence is shortened. Albert and Altshuler use the log-normal distribution to represent time to tumor occurrence assuming the standard deviation to be independent of dosage.

Dr. Nancy Mann, in addition to lively introductions and elevation of the esthetic level of the speakers' platform, discussed the Weibull distribution. The Weibull distribution for time to tumors has been suggested by human cancers, (Cook, Doll, and Fellingham 1969; Lee and O'Neill, 1971); $I = bd^m(t\text{-}w)^k$, where I is the incidence rate of tumors at time t, b is a constant depending on experimental conditions, d is dosage, w (the minimum time to the occurrence of an observable tumor), m and k are parameters to be estimated. Also, Day (1967), Peto, Lee and Paige (1972) and Peto and Lee (1973) have considered the Weibull distribution for time to tumor occurrence. Theoretical models of carcinogenesis also predict the Weibull distribution (Pike, 1966). Theoretical arguments and some experimental data suggest the Weibull distribution where tumor incidence is a polynomial in dose times a function of age.

The log-normal distribution of tumor times corresponds to the probit transformation as employed in the Mantel-Bryan procedure. Use of the Weibull distribution for time to tumor leads to an extreme value distribution relating tumor response to dosage (Chand and Hoel, 1973). Hoel (1972) gives techniques when adjustments must be made for competing causes of death. Albert and Altshuler (1973) discuss other distributions of time to tumor. What must be done is to encourage more experimentation and statistical research on survival studies. It probably would be much more palatable to set safe doses on the basis that the probability of a tumor is remote if an animal lived to, say, twice its normal lifespan; rather than to say that the probability is remote that an animal develops a tumor during its lifespan. However, it is still the latter quantity which is of concern. We have invited Doctors Albert and Altshuler to examine the data bases at NCTR in hope of increasing the documentation of their model in large carefully controlled animal experiments.

Variation in Exposure

I am borrowing Doctor Newill's slide to demonstrate variation in exposure. Human intake of a chemical varies among individuals and varies daily for a given individual. The simplest approach and perhaps adequate for our current state of knowledge is to calculate risks for anticipated "maximum" exposure levels. This gives additional conservatism for any prediction technique. Some contend that it is not necessary to attempt to protect every last individual with unusual habits, but to base predictions on average intake, for example, in an attempt to estimate the actual risk to the population.

Mathematically, the proper approach to calculate the risk for the total population is to calculate the risk for a given dosage (a most difficult task as discussed in the previous sections) and then to multiply that risk by the proportion of the time that dosage occurs followed by integration over the distribution of dosages. Generally, the distribution of dosages for an environmental chemical in a human population would be unknown. Thus, introducing variation in consumption adds another dimension to be investigated to an already complicated problem. This, in effect, gives the average risk and does not consider a segment of the population which may be at high risk. Indeed the lively discussion from the floor about NO_x levels in a kitchen speaks to this problem.

I wish to share with you an outline of one of the chronic dose-response studies being conducted with 2-AAF, a carcinogen, at NCTR. I feel that an introduction into the needs of FDA and EPA may be of value in understanding our approach.

1. Responsibilities

A. *Food and Drug Administration.*— The general public may be exposed to chemical carcinogens from the food it consumes, the drugs it uses, and the cosmetics it enjoys. It is the responsibility of the FDA to determine the risk involved in the use of these commodities by the general public. However, it must be recognized that there is no way to guarantee absolute safety. Small populations of experimental subjects, either animal or man, provide an imprecise basis for comparison to a large human population of variable genetic and disease states, cultural backgrounds, and ages. Furthermore, toxicological assessments are made on individual chemicals and humans are exposed to a milieu of interacting chemicals. Leading us somewhat out of this apparently chaotic situation, studies with laboratory animals have shown that nearly all chemicals that are carcinogenic in man are also carcinogenic in one or more animal species although the tumors may be of a different type. Thus, the carcinogenic properties of a chemical may be detected in experimental animals just as we detect the life shortening or lethal properties of a chemical. It is apparent that, while cancer testing in ani-

mals in terms of "an all-or-none effect" has reached a sophisticated level of development, the area of quantitative dose-response relationship for testing of chemical carcinogens needs much more in-depth study. It is precisely this area that must develop if a rational assessment is to be made of "acceptable risk" and "acceptable daily intake" of a chemical carcinogen in our food, drugs, or cosmetics.

B. *Environmental Protection Agency.*—An understanding of why certain environmental agents produce adverse effects and the circumstances that determine the severity of these effects is the basis of all environmental health regulatory control. Carcinogenic substances pose a hazard to man and the environment through several distinct pathways. The most obvious of these is direct ingestion. Control of chemical food additives is the responsibility of the Food and Drug Administration as indicated earlier. However, the presence of unintentional residues in food such as residues of pesticides and other toxic substance is a responsibility of the Environmental Protection Agency. This responsibility extends to cover chemicals present in water and air and their effects not only on man but to any component of the environment.

In terms of man, it is clear that the main chemical carcinogenic hazards result from exposure associated with food, water, and air. The degree of the hazards involved depends on many factors. One of the basic factors involves the quantitative aspects of a chemical carcinogenic action. Although the many factors involved are important, it is the quantitative aspects which has not received intensive study and which is basic to determining acceptable daily exposure levels for a chemical carcinogen. These acceptable daily exposure levels, in turn, are a basic requirement to setting standards for chemical carcinogens in our food and our environment.

C. *Others.*—The National Institute of Occupational Safety and Health (NIOSH) of the Department of Health, Education, and Welfare and the Occupational Safety and Health Administration (OSHA) of the Department of Labor share the regulatory responsibilities for the control of hazards in the occupational environment. These responsibilities involve the control of hazards associated with occupational exposure to chemical carcinogens. Thus, the concepts methodology involved in evaluating quantitatively the risk involved in exposure to these substances is a basic need of these agencies.

Finally, it must be recognized that the necessity of information on the quantitative determination of acceptable exposure levels to carcinogens is basic not only to the regulatory agencies, but is of particular importance to those involved in meeting our extensive chemical needs. Chemicals as part of various products, drugs, pesticides, food additives, water additives, etc., are a necessity of life and, in turn, create a necessity of information permitting their use within acceptable limits of risk.

II. *Deficiencies in Work to Date and Factors to be Considered in Protocol Development.*—Most of the deficiencies in carcinogenic testing result mainly from the concept that this testing involves only the determination as to whether or not a compound can be made to produce a neoplastic tumor. However, it is now recognized that carcinogenic testing must, of necessity, consider both qualitative and quantitative factors. The main deficiencies in past studies of these factors involve primarily two areas, i.e., experimental design and definition of end points.

A. *Experimental Design:*

1. *Statistics.*—The bulk of the technical literature reflects the lack of statistically valid experimental design including adequate numbers of animals at low levels of carcinogenic response.

2. *Dose Response.*—Only limited use of dose response studies are reported

in the technical literature for the purpose of determining tumor incidence and time to tumor in terms of dose rate and total dose. The prediction of risk at a given exposure level requires dose response information.

3. *Low-Dose Studies.*—Little information is available on dose-response studies at low levels of exposure and response. At low levels of exposure environmental factors may alter extensively the quantitative aspects of a response.

4. *Mathematical Models.*—There is only limited mathematical definition of the dose-response curves at low levels of exposure in terms of variables affecting a chemical carcinogenic response.

5. *Life-Shortening.*—There is limited use of experimental designs which permit proper observation and evaluation of life-shortening effects of chemical carcinogens in relation to dosage.

6. *Age Sensitivity.*—The hazards involved in exposure to a chemical carcinogen depend not only on the nature of the chemical itself, the route of exposure, and the extent of exposure in terms of amount of time, but also on the susceptibility of the animal at the time of exposure. There are only limited studies available on the influence of age on the sensitivity of an animal to a chemical carcinogen.

7. *Recovery.*—There is a lack of evaluation of the possible regression or progression of pretumorous lesions such as hyperplasia in relation to dosage.

8. *Tumor Growth Rate.*—The technical literature shows an impressive lack of study of the possible dependency of tumor growth rate on dosage.

9. *Reproducibility of Results.*—There is an extensive lack of evaluation of the quantitative reproducibility of chemical carcinogenic testing.

B. *Endpoints:*

1. *Tumorigenesis.*—Tumorigenesis is as important an endpoint as carcinogenesis. Benign tumors may cause death in man and animals without ever undergoing malignant transformation. There can be no doubt from a survey of the technical literature that benign neoplasms are often precursors of malignancies. In the light of present knowledge, all tumorigens must be regarded as potential carcinogens. Hyperplasia and number, type, grade, and individual distribution of tumors must all be carefully used as endpoints in the evaluation of chemical carcinogenesis.

2. *Time to Tumors.*—In some cases the only manifestation of an effect consists of an earlier occurrence of tumors in the treated animals than in the controls. Time to tumor may be a very sensitive endpoint permitting estimation of "acceptable exposure levels" from dose-time to tumor curves. This endpoint in chemical carcinogen testing merits further in-depth study.

3. *Life Shortening.*—As indicated earlier, there is limited use of experimental design which permit proper observations and evaluation of life-shortening effects of chemical carcinogens in relation to dosage.

4. *Pathology.*—It is of the utmost importance that a complete and accurate pathological examination be conducted on all animals used in carcinogenesis studies. There is no doubt that benign tumors may cause death without undergoing malignant transformation. All lesions, including precancerous lesions such as hyperplasia, must be described. Number, type, grade, and individual distribution of tumors must all be carefully evaluated in a chemical carcinogenesis study. The lack of proper pathological capabilities often limits this most critical aspect of such a study.

5. *Biochemistry.*—The evaluation of carcinogenic hazards for man is based on a judgment of all available information. That is, it is based not only on the carcinogenic bioassay, toxicity tests, epidemiological data, and on the extent and route of exposure of man, but also on

metabolic, biochemical, and pharmacokinetic studies. Each compound must be evaluated individually on the nature of its absorption, distribution, metabolism, retention, and excretion.

Ancillary support experiments, generally independent of the large ED_{01} Barrier Study, will be undertaken within the programs of the Divisions of Chemistry and Comparative Pharmacology. The purpose of these studies will be to define appropriate biochemical endpoints and the role of pharmacokinetics to aid in the evaluation and interpretation of the large carcinogenic bioassay. For the most part, these studies will be undertaken with mice not maintained in the barrier experiment. A select number of biochemical parameters, however, can be measured in some mice at the time that these animals are removed from the ED_{01} experiment.

Biochemical endpoints, as an indicator of or response to carcinogen exposures, are not usually included in carcinogenic bioassays. Identification of reliable indices that relate directly to tumorigenesis would be invaluable to possibly define susceptible or non-susceptible individuals in an animal population or to possibly determine the time to onset of irreversible lesions during a precancerous induction period. This concept is highly important to and related to the chronic low dose carcinogenic bioassays. However, the current status of this concept has not been definitely proven or confirmed and, as such, must be considered as an activity peripheral to the large bioassay study at this time.

Logically, any stimulus such as a chemical carcinogen producing an anabolic or precancerous change in a tissue such as liver should produce some response, such as stimulation or inhibition of an enzyme(s) that can be detected biochemically in the affected tissue or possibly in the blood. The inherent problem is to select or find the proper biochemical endpoint. Several prospective endpoints have been defined, but their potential utility as predictors or indicators of response remains to be established. Research activities in some of these biochemical indicators are centered in the Divisions of Chemistry and Comparative Pharmacology.

The role of DNA repair in the toxic and carcinogenic effect of 2-AAF also must be considered in relation to the chronic low dose bioassay. Recent evidence resulting from studies on the effect of radiation on biological systems indicates that mammalian cells have the capability of repairing damage to their DNA. More recently it has been demonstrated that many chemicals such as AAF form covalent bonds with DNA and are removed by a process of "unscheduled DNA synthesis" or DNA "repair synthesis". The process appears to involve the excision of the damaged segment of DNA with concomitant replacement by repair synthesis. The importance of this process was made evident with the demonstration that the resistance of numerous tumors to chemotherapeutic agents could be correlated with their level of DNA repair activity. Tumors resistant to chemotherapeutic agents were found to be susceptible in the presence of DNA repair inhibitors such as caffeine or chloroquine.

It is often assumed that DNA repair always acts in a protective way by removing damaged DNA segments or bound chemical residues. It is known, however, that the probability of an error in DNA replication which might result in a mutation increases with the extent of DNA synthesis. The possibility of AAF producing mutations in DNA by stimulating extensive DNA repair synthesis is a real one and must be considered in any study concerning the role of DNA repair in carcinogenesis. In any event, a more complete understanding of how a cell repairs the damage inflicted upon its genetic information by chemicals in general and carcinogens in particular will be necessary. An understanding of the role of DNA repair in carcinogenesis is basic to the question of whether small chemical insults to a cell are completely repaired or accumulate over a long period of chronic exposure.

A study to investigate the role of DNA repair in the carcinogenic process has been initiated within the Division of Comparative Pharmacology. The specific objectives of this study are to seek a correlation between DNA repair and tumorigenesis under several experimental conditions, to determine the effects of acute and chronic doses of 2-AAF on DNA repair, to evaluate the effect of DNA repair inhibitors on tumor incidence, and to investigate the possible interrelationships between DNA repair and cell division in carcinogenesis. Results from this study will provide valuable input to the understanding and interpretation of the chronic study with 2-AAF.

6. *Pharmacokinetics.*—In order to provide a firmer basis for evaluation of results obtained in the large chronic low-dose carcinogenic bioassay, it will be essential to develop a correlation between dietary level of the carcinogen, total and/or daily intake of chemical, incorporation of chemical into the target site (bladder in this instance) and the incidence of bladder tumors as a function of duration and level of exposure. Involved also in this correlation is the need to evaluate the role of blood levels (total as well as unbound) and urinary excretion patterns of the chemical and/or its metabolites. The overall concept or process described is the basis and definition of pharmacokinetics. Pharmacokinetics basically measures rates of chemical absorption, distribution, tissue binding and storage, metabolism, and elimination. Elimination in this case meaning excretion through urine, feces, and expired air. Mathematical models are designed to analyze results by means of computer simulation.

With 2-AAF, a unique opportunity is presented to relate dietary levels and feed consumpation to relative levels of the compound or metabolites in blood, urine, urinary bladder, and incidence of bladder tumors. The key comparisons will have to evolve based on chronic exposure of the animal to the test compound. However, to develop a model on which to evaluate results from chronic exposure, it was necessary to undertake a series of acute and subacute experiments designed to determine as a function of dose level the absorption, distribution, metabolism, excretion, and bladder binding of 2-AAF following single and multiple P.O., I.P., and I.V. doses of chemical as well as following dietary exposure. Based on models developed from these studies, responses were predicted for chronic exposure to 2-AAF; the accuracy of these predictions will be verified from results that will be obtained in a chronic exposure metabolism study. In terms of the large ED_{01}-2-AAF study, this approach will be limited to establishing dosage levels of 2-AAF to concentrations of the compound and/or its metabolites in blood with reference to time and to the effect on the endpoint being studied.

The more in-depth pharmacokinetic study will be undertaken within the Division of Comparative Pharmacology as a separate study from the large ED_{01}—2-AAF experiment.

III. *Approaches.*—It is clear that human exposure to many chemical carcinogens is inevitable at the present time and in the foreseeable future. It follows that a need exists for capabilities which would permit an evaluation of the relative hazards posed by different chemical carcinogens. The development of methodology for adequately evaluating carcinogenic risk involves two major approaches. The first is the establishment of a carcinogen dose-response relationship using various endpoints such as tumor prevalence, time to tumor, life shortening, etc. This carcinogen dose-response relationship must permit some mathematical extrapolation downward on the curve so as to facilitate determination of risk at levels of realistic exposure. These are the primary objectives of this study. The second approach which is beyond the scope of this study is to develop methodology and concepts which will permit extrapolation of results to man.

A. *Dose Response.*—The dose-response of tumor prevalence in terms of dose rate and total dose, giving appropriate consideration to cause of death, will be determined. Such data should rival the best mathematical model for the conservative extrapolation from dose-tumor prevalence curves to an exposure level that would pose a "socially acceptable risk".

B. *Time to Tumor.*—There is increasing interest in the time to tumor dose-response relationship in chronic studies. Early tumors have much more impact on lifespan than do late tumors. Furthermore, for some carcinogens, in particular at low levels of exposure, the only manifestation of an effect consists of an earlier occurrence of tumors in the treated animals than in the controls, the tumor prevalence being the same in both. The prevalance of tumors as a function of age (time to tumor) over the life span of the animals provides a better description of the tumorigenic process than at a single point in time of sacrifice or the total prevalence of tumors over the life span. Two problems need to be resolved. First, the relationship between dosage and median time to tumor must be established. Secondly, given the dosage, the distribution of time to tumors must be established to estimate the prevalence of tumors during the life span for a given dosage. The survival group (life span group) in the experimental design of this protocol will provide good data for such analyses.

C. *Life Shortening.*—The lifespan portion of the experimental design allows the evaluation of life shortening as an endpoint for chronic studies. All of the information gathered on time to tumor development also demonstrate life shortening for lethal tumors.

D. *Age Sensitivity.*—Hazards involved in exposure to a chemical carcinogen depend not only on the nature of the chemical itself, the route of exposure, and the extent of exposure in terms of amount and time, but also on the susceptibility of the animals at the time of exposure. Age sensitivity studies are being conducted in a separate experiment in order to permit evaluation of possible age sensitivity to 2-AAF in relation to specific periods of treatment of mice with this compound during serial sacrifice and serial treatment phase of the ED_{01}—2-AAF Study.

E. *Regression or Progression of Effects.*—The experimental design permits groups fed 2-AAF for 6, 9, and 12 months and sacrificed to be compared with groups fed the same length of time but sacrificed at 18 months. The purpose is to study the possible regression or progression of pretumorous lesions such as hyperplasia in relation to dosage and time.

F. *Tumor Growth Rate.*—The experimental design permits an approach to the question of the possible dependence of tumor growth rate on dosage and treatment as will be revealed in the serial sacrifice and serial treatment phases of the study.

IV. *Experimental Design.*—The design contains both basic types of experiments: survival (lifespan) and serial sacrifice. The serial sacrifice portion is sub-divided into continuous and discontinued feeding.

A. *Pilot Studies.*—A pilot study in which 2-AAF was administered in the feed of mice for eighteen months established the suitability of the strain and sex and gave indications of the dosage-time range to be used in the more extensive ED_{01} Study. The results of the pilot study will be published elsewhere.

B. *Animals*

1. *Species, Strain, and Sex.*—Mice were selected for this study because of the need for large numbers of animals required for the statistical validity of dose-response studies at low levels of exposure (dosage) to a chemical carcinogen. The choice of mice is further substantiated if one considers the availability of well defined inbred strains of animals having a relatively short life span.

The BALB/c strain was selected because of the lack of spontaneous bladder tumors in contrast to its high susceptibility to 2-AAF induction of these tumors. Based on general concepts and on Pilot Study results, the main objective, that is the development of suitable mathematical description of a chemical carcinogen dose-response curve permitting extrapolation from high to low levels of response, was determined to be equally possible with either sex of the selected strain. The dosage range studied in the pilot experiment gave better data points on the curve for the female than for the male BALB/c mice, therefore, females were selected for the ED_{01} Study. The nature of the dose-response curve at low levels of prolonged exposure to a chemical carcinogen could be studied using both sexes of several strains of mice and using several carcinogens and types of tumors. Such an extensive experiment would be considered best after a limited and more circumscribed study revealed the need, amplified the approach and indicated the success and usefulness of such an undertaking.

2. *Age of Animals.*—All animals allocated to the experiment will be weanlings, three to four weeks of age.

3. *SPF-DF Animals.*—All mice used in the experiment will be "specific pathogen free defined flora" animals derived in the breeding colony of NCTR from a Charles River substrain of BALB/c mice.

C. *Dosages.*—Based on the Pilot Study results, seven dosages expressed as ppm of 2-AAF in the feed were selected to give approximately a tumor prevalence of 64 through 1% as indicated in Table 5. It must be recognized that the dose-response relationship expressed in Table 5 is based on an 18-month study in which the animals used were not SPF-DF animals maintained under barrier conditions. Furthermore, although the mice were BALB/c females, they were from a commercial source and were not derived from the NCTR mice breeding colony. It must be stressed that the dose response relationship is considered the best approximation which could be made when all the available data were considered. The accuracy of this approximation can only be determined by the ED_{01}-2-AAF Chronic Study.

Table 5.—Bladder tumor prevalence with 2-AAF in feed.

2-AAF Concentration in Feed (ppm)	Bladder Tumor Prevalence (ED%)
200	64
175	32
150	16
100	8
50	4
25	2
10	1

D. *Type and Duration of Treatments.*—The survival phase of the study involves a lifespan exposure to 2-AAF in the feed. The animals in this phase of the study will be removed from the experiment as they become moribund. The serial sacrifice phase involves treatment for and sacrifice at 6, 7, 8, 9, 12, 15, and 18 months. The serial treatment or recovery study involves treatment for 6, 9, and 12 months followed by recovery and sacrifice at the eighteenth month of entering the experiment.

E. *Statistics.*

1. *Grouping and Randomization of Animals.*—As animals are received from Animal Husbandry Division, they will be randomly allocated to the various experimental groups. This will insure that any differences in animals, feed, laboratory conditions, or handling will be approximately the same for all experimental groups. For ease of operation, treatments will be grouped in tiers of six cages on a rack. This will also average out floor-to-ceiling differences in temperature, light, and humidity if these should be important factors.

2. *Sequential Entry.*—The animals will be placed on experiment, a room at a

time. Each barrier room pertinent to the ED_{01}-2-AAF Chronic Study will be loaded at the rate of two racks per week, requiring seven weeks to load a room. The randomization of animals to treatments described above should nearly eliminate most changes that occur with time.

3. *Replication of Module.*—To conduct the experiment, the module presented in the experimental design will be replicated six times. That is, the experiment will be conducted in six barrier rooms. This should provide adequate numbers of animals to estimate dose-response slopes within ±50% and to estimate the ED_{01} levels within a factor of two. Thus, mathematical models that differ by a factor of three at the ED_{01} levels can be detected.

Table 6 presents the number of different dose groups and the number of different experimental components in each of the six rooms.

Summary

Many facets of life, including food products currently consumed involve risks. It is a worthy goal to strive for absolute safety, but it is impossible to demonstrate absolute safety experimentally.

The problem is not to determine whether or not a socially necessary compound is a carcinogen at high experimental doses, but to estimate risks at low dosages approximating human exposure levels. Such estimation procedures should require the setting of tolerances based on the certainty of experimental results.

In order to observe biological effects with adequate statistical precision from a reasonable number of animals, experimental dosages are generally well above human exposure levels. Thus, extrapolation of effects to lower dosages must be made to estimate risks.

Estimated risks vary widely depending on the mathematical model used for extrapolation and the values of the parameters used in the model.

Age specific tumor rates may give an incomplete description of the tumorigenic process. More emphasis is needed on survival studies in which time to tumor occurrence is studied. A parameter such as life shortening may be more meaningful than the proportion of animals developing tumors.

More information is needed on comparisons of results for various chemicals and species from survival studies, including the effect of dosage on the param-

Table 6.—2-AAF Chronic Study (Cages/Room).

Purpose	Survival	Serial Sacrifice							Serial Treatment			Diagnostic	Total
Mo. of Sacrifice	None	18	15	12	9	8	7	6	18	18	18		
Mo. on 2-AAF	Life	0-18	0-15	0-12	0-9	0-8	0-7	0-6	0-12	0-9	0-6		
ED_{64}	6	12	6	6	6	6	6	6	6	6	6	6	78
ED_{32}	6	12	6	6	6	6	6	6	6	6	6		72
ED_{16}	12	24	12	12	6	6	6	6	12	12	12		120
ED_8	12	24	18	18	12	12	12	12	18	18	18		174
ED_4	18	36	24	18									96
ED_2	36	72	36										144
ED_1	72	144											216
ED_0	18	36	12	12	6	6	6	6				6	108
Total	180	360	114	72	36	36	36	36	42	42	42	42	1008

Months denoted as time on treatment.
Dosages based on expected % of bladder tumors at 18 months.
Four mice per cage; 72 cages per rack; 14 racks per room.
Repeat experiment in 6 rooms for a total of 24,192 animals.
Replace cages for diagnostics as animals are depleted after 6 month serial sacrifice.
Use BALB/c female weanling mice.

eters of the time pattern response both in man and animals.

More information is needed on human intake of various chemicals so that estimates of risk can take into account variation in exposure rather than calculating risks for average or "maximum" consumption.

I would be remiss if I did not stress that the major advantage of animal toxicology over human epidemiology is that the toxicity can be predicted *before* human exposure.

I hope that I have adequately dealt with some of the questions from the floor and discussed some of the needs in statistics from the vantage point of a toxicologist. I have attempted to identify one of the scientific programs at NCTR, the chronic low level dose response experiment with 2-AAF.

I must emphasize that there are two other areas in toxicology equally needy in basic dose response experimentation: mutagenesis and teratogenesis. The NCTR is launching programs in these areas equal in depth to the chronic study I have described. In closing let me take a few minutes of your time to discuss why we cannot make decisions as to the adverse health of any chemical in a vacuum. We are now facing a severe fossil fuel energy shortage. Most of the world is facing a severe nutritional shortage. Consider the following chain of events. The United States exports grain and improves our balance of payments. The United States imports oil and shifts our balance of payments toward a deficit. The EPA wishes to improve air quality and among several approaches is the use of low sulfur fuel and control technology. Both cost money. An energy crunch comes along with escalating costs. The FDA, also operating under laws to protect the public from adverse health effects, limits the use of growth promotants; the EPA controls the use of certain pesticides and the result is less grain available after domestic use. The use of natural gas is limited in the production of fertilizers and farmers may have less fuel. All of this results in less grain at a higher cost. When less grain is available for export we have less money for low sulfur fuels and the air becomes less clean or control technology costs escalate. The point is that agricultural production and many other components of a highly technological society are very closely webbed with options for a clean environment. More attention should be placed on legislation consistent with integrated control and quality. It is the task of the toxicologist and statistician to provide the decision makers with data which can be used in establishing relative health effects.

References Cited

Albert, R. E., and Altshuler, B. 1973. Considerations relating to the formulation of limits for unavoidable population exposures to environmental carcinogens. In *Radionuclide Carcinogenesis* by Ballou, J. E. *et al* editors, AEC Symposium Series, CONF-72050, NTIS, Springfield Va., 233–253.

Altshuler, B. 1970. Theory for the measurement of competing risks in animal experiments. Math. Biosciences 6: 1–11.

Berenblum, J., and Haran, N. 1955. The influence of dose of a carcinogen, emptiness of stomach and other factors on tumor induction in the fore stomach of mouse. Cancer Res. 15: 504–509.

Berg, J. W. 1964. Disease-oriented end results. Cancer 17: 693–707.

Berkson, J., and Elveback, L. 1960. Competing exponential risks with particular reference to smoking and lung cancer. J. Amer. Stat. Assoc. 55: 415–28.

Blum, H. F. 1959. *Carcinogenesis by Ultraviolet Light.* Princeton University Press, Princeton, N. J.

Breslow, N. 1970. A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. Biometrika 57: 579–94.

Chand, N., and Hoel, D. G. 1973. A comparison of models for determining safe levels of environmental agents. Conference on Reliability and Biometry, Florida State University, Tallahassee.

Chiang, C. L. 1968. *Introduction to Stochastic Processes in Biostatistics.* Wiley, N. Y.

Cook, P. J., Doll, R., and Fellingham, S. A. 1969. A mathematical model for the age distribution of cancer in man. Int. J. Cancer 4: 93–112.

Culter, S. J., and Ederer, F. 1958. Maximum utilization of the life table method in analyzing survival. J. Chron. Dis. 8: 699–712.

Day, T. D. 1967. Carcinogenic action of cigarette smoke condensate on mouse skin. Br. J. Cancer 21: 56–81.

Druckrey, H. 1967. Quantitative aspects of chemical carcinogenesis. U.I.C.C. Monograph Series, Vol. 7, Potential Carcinogenic Hazards from

Drugs, (Evaluation of Risks) pp. 60–78. Editor Reńe Truhaut. Springer-Verlag, N. Y.

FDA Advisory Committee on Protocols for Safety Evaluation. 1971. Panel on carcinogenesis report on cancer testing in the safety evaluation of food additives and pesticides. Tox. Appl. Pharm. 20: 419–438.

Friedman, Leo. 1973. Personal communication.

Gross, M. A., Fitzhugh, O. G., and Mantel, N. 1970. Evaluation of safety for food additives: An illustration involving the influence of methyl salicylate on rat reproduction. Biometrics 26: 181–194.

Hoel, D. G. 1972. A representation of mortality data by competing risks. Biometrics 28:475–488.

Hoel, D. G. and Walburg, H. E. 1972. Statistical analysis of survival experiments. J. Nat. Cancer Inst. 49: 361–372.

Kaplan, E. L. and Meier, P. 1958. Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53: 457–81.

Kimball, A. W. 1958. Disease incidence estimation in populations subject to multiple causes of death. Bull. Int. Stat. Inst. 36: 193–204.

Lee, P. N., and O'Neill, J. A. 1971. The effect both of time and dose applied on tumor incidence rate in benzopyrene skin painting experiments. Br. J. Cancer 25: 759–770.

Mantel, N., and Bryan, W. R. 1961. Safety testing of carcinogenic agents. J. Nat. Cancer Inst. 27: 455–470.

Mantel, N., and Haenszel, W. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. J. Nat. Cancer Inst. 22: 719–48.

Moeschberger, M. L., and David, H. A. 1971. Life-tests under competing causes of failure and the theory of competing risks. Biometrics 27: 909–33.

Murray, J. L., and Axtell, L. M. Impact of cancer: Years of life lost due to cancer mortality (to be published in the J. Nat. Cancer Inst.).

Peto, R., Lee, P. N., and Paige, W. S. 1972. Statistical analysis of the bioassay of continuous carcinogens. Br. J. Cancer 26: 258–261.

Peto, R., and Lee, P. 1973. Weibull distributions for continuous-carcinogenesis experiments. Biometrics 29: 457–470.

Peto, R., and Pike, M. C. 1973. Conservatism of the approximation $\Sigma(O-E)^2/E$ in the logrank test for survival data or tumor incidence data. Biometrics 29: 579–589.

Pike, M. C. 1966. A method of analysis of a certain class of experiments in carcinogenesis. Biometrics 22: 142–161.

Sampford, M. R. 1952. The estimation of response-time distributions II. Multi-stimulus distributions. Biometrics 8: 307–53.

Summary Address for the Symposium— Statistics and the Environment

Fred C. Leone

Executive Director, American Statistical Association, 806 15th St., N.W., Washington, D. C. 20005

As the pilot said to his passengers while trying to find his way along the coast in inclement weather, "Folks, I have some good news and some bad news", I can say that I have some good and some bad to report. But first the bad, then the good.

The objective of this Symposium as stated in the program was "to provide a forum for the interchange of ideas of mutual interest among experts in toxicology and environmental areas with specialists in the statistical techniques of data gathering and analysis. This is not a meeting where statisticians will speak statistically to their colleagues, or environmentalists will converse in their own language to their co-scientists. It is hoped that attempts to solve environmental problems will be enhanced by an interdisciplinary approach resulting from the communication among the pertinent professions."

If in fact this meant that we will solve many problems here, we have failed and failed miserably. If the purpose is, as